wwPDB X-ray Structure Validation Summary Report (i)

May 21, 2020 – 11:12 pm BST

PDB ID	:	3WXI
Title	:	Crystal structure of trypanosoma brucei gambiense glycerol kinase (ligand-free form)
Authors	:	Balogun, E.O.; Inaoka, D.K.; Shiba, T.; Kido, Y.; Tsuge, T.; Nara, T.; Aoki, T.; Honma, T.; Tanaka, A.; Inoue, M.; Matsuoka, S.; Michels, P.A.M.; Kita, K.; Harada, S.
Deposited on	:	2014-08-01
Resolution	:	2.90 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

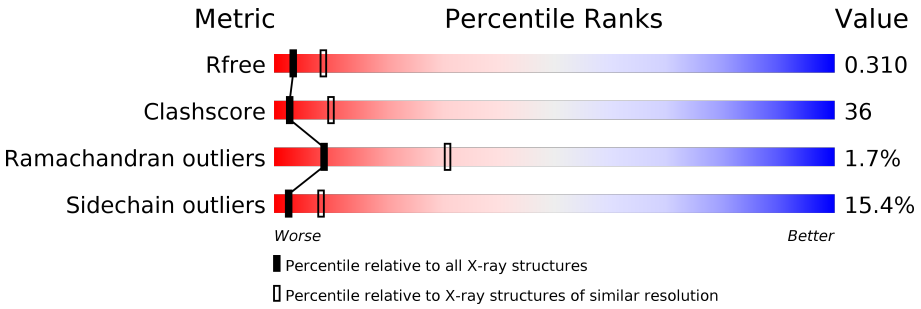
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5%

Mol	Chain	Length	Quality of chain
1	A	518	63% 29% 7%
1	B	518	53% 37% 9%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7906 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Glycerol kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	513	Total	C	N	O	S	0	0	0
			3957	2499	694	731	33			
1	B	512	Total	C	N	O	S	0	0	0
			3940	2486	692	729	33			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	GLY	-	EXPRESSION TAG	UNP D3KVM3
A	-4	ILE	-	EXPRESSION TAG	UNP D3KVM3
A	-3	ASP	-	EXPRESSION TAG	UNP D3KVM3
A	-2	PRO	-	EXPRESSION TAG	UNP D3KVM3
A	-1	PHE	-	EXPRESSION TAG	UNP D3KVM3
A	0	THR	-	EXPRESSION TAG	UNP D3KVM3
B	-5	GLY	-	EXPRESSION TAG	UNP D3KVM3
B	-4	ILE	-	EXPRESSION TAG	UNP D3KVM3
B	-3	ASP	-	EXPRESSION TAG	UNP D3KVM3
B	-2	PRO	-	EXPRESSION TAG	UNP D3KVM3
B	-1	PHE	-	EXPRESSION TAG	UNP D3KVM3
B	0	THR	-	EXPRESSION TAG	UNP D3KVM3

- Molecule 2 is water.

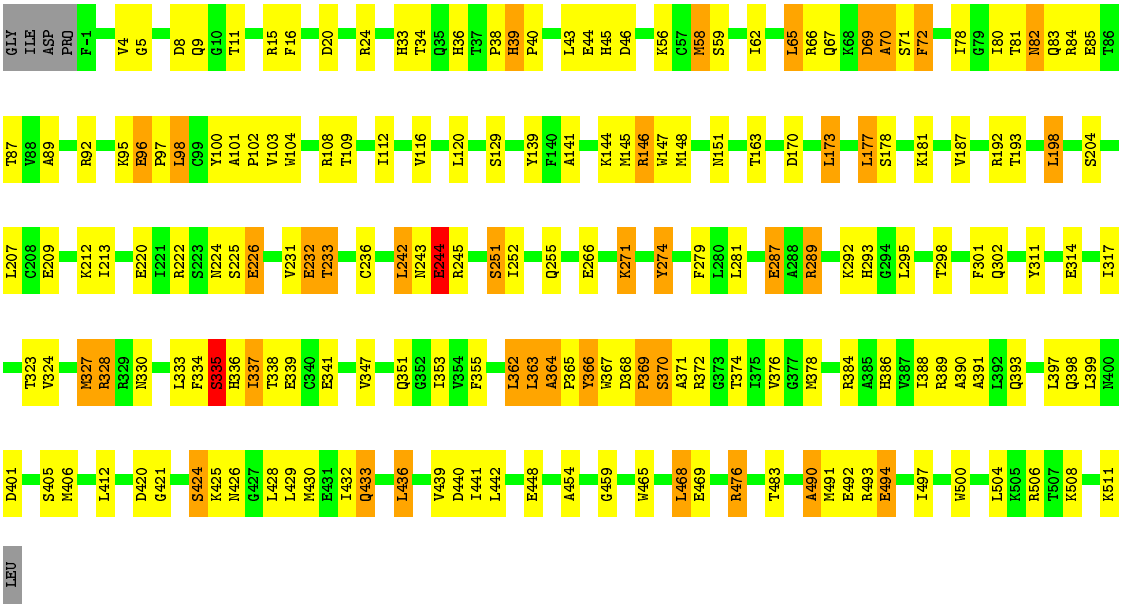
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	O	0	0
			2	2		
2	B	7	Total	O	0	0
			7	7		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

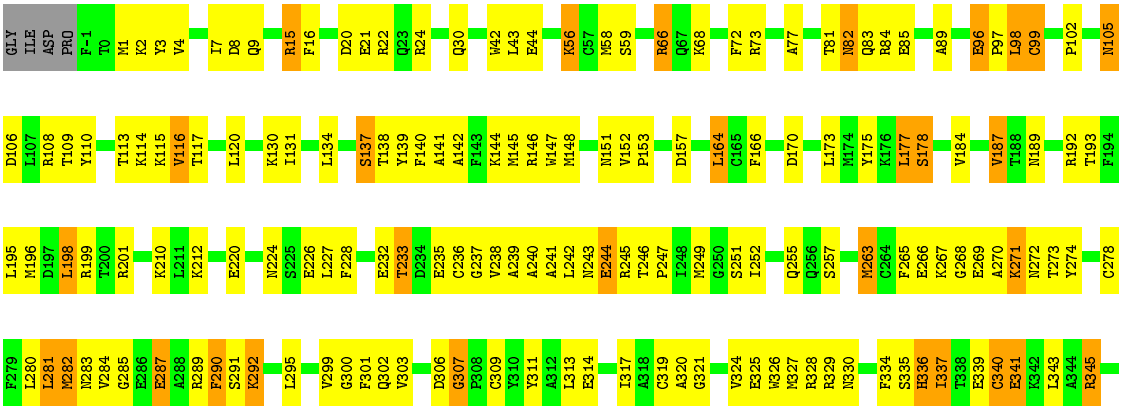
- Molecule 1: Glycerol kinase

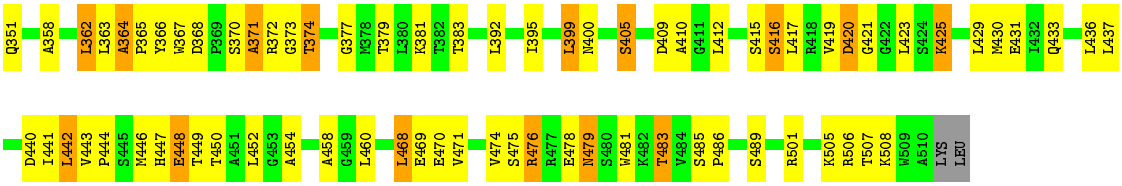
Chain A:



- Molecule 1: Glycerol kinase

Chain B:




Q351
A358
L362
L363
A364
P365
Y366
W367
D368
P369
S370
A371
R372
G373
T374
G377
M378
T379
L380
K381
T382
T383
L392
I395
L399
N400
S405
D409
A410
G411
L412
S415
S416
L417
R418
V419
D420
G421
G422
L423
S424
K425
L429
M430
E431
I432
Q433
L436
L437
D440
I441
L442
V443
P444
S445
M446
H447
E448
T449
T450
A451
L452
G453
A454
A458
G459
L460
L468
E469
E470
V471
V474
S475
R476
R477
E478
N479
S480
W481
K482
T483
V484
S485
P486
S489
R501
K505
R506
T507
K508
W509
A510
LYS
LEU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	68.76Å 131.98Å 148.61Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.90 29.75 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.5 (30.00-2.90) 99.5 (29.75-2.90)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.24 (at 2.90Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.202 , 0.282 0.253 , 0.310	Depositor DCC
R_{free} test set	1539 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	101.7	Xtriage
Anisotropy	0.053	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 35.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	7906	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.70% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.77	0/4039	0.85	2/5465 (0.0%)
1	B	0.67	1/4021 (0.0%)	0.76	0/5442
All	All	0.72	1/8060 (0.0%)	0.81	2/10907 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	307	GLY	N-CA	5.10	1.53	1.46

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	369	PRO	CA-N-CD	-8.29	99.90	111.50
1	A	70	ALA	N-CA-C	-5.51	96.12	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3957	0	3968	265	0
1	B	3940	0	3943	304	0
2	A	2	0	0	0	0
2	B	7	0	0	0	0
All	All	7906	0	7911	563	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 36.

The worst 5 of 563 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:238:VAL:CG1	1:B:240:ALA:HB3	1.44	1.43
1:B:481:TRP:HZ3	1:B:483:THR:CG2	1.33	1.40
1:A:364:ALA:HB3	1:A:366:TYR:CE2	1.56	1.40
1:B:238:VAL:CG2	1:B:240:ALA:H	1.46	1.28
1:B:238:VAL:HG11	1:B:240:ALA:CB	1.65	1.27

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	511/518 (99%)	465 (91%)	39 (8%)	7 (1%)	11	36
1	B	510/518 (98%)	456 (89%)	44 (9%)	10 (2%)	7	27
All	All	1021/1036 (99%)	921 (90%)	83 (8%)	17 (2%)	9	31

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	335	SER
1	A	337	ILE
1	A	226	GLU
1	A	244	GLU
1	B	340	CYS

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	421/425 (99%)	360 (86%)	61 (14%)	3 9
1	B	418/425 (98%)	350 (84%)	68 (16%)	2 7
All	All	839/850 (99%)	710 (85%)	129 (15%)	2 8

5 of 129 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	506	ARG
1	B	131	ILE
1	B	468	LEU
1	A	511	LYS
1	B	96	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	413	ASN
1	B	23	GLN
1	B	243	ASN
1	A	386	HIS
1	B	293	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.